

0590
0517

46



OIPE

ENTERED

RAW SEQUENCE LISTING

DATE: 05/13/2002

PATENT APPLICATION: US/10/026,994

TIME: 09:38:32

Input Set : A:\GC698-SEQLIST.txt

Output Set: N:\CRF3\05102002\J026994.raw

4 <110> APPLICANT: Dunn-Coleman, Nigel
 5 Goedegebuur, Frits
 6 Ward, Michael
 7 Yao, Jian
 9 <120> TITLE OF INVENTION: EGVI Endoglucanase and Nucleic Acids
 10 Encoding the Same
 12 <130> FILE REFERENCE: GC698
 14 <140> CURRENT APPLICATION NUMBER: US 10/026,994
 C--> 15 <141> CURRENT FILING DATE: 2002-04-30
 17 <160> NUMBER OF SEQ ID NOS: 4
 19 <170> SOFTWARE: FastSEQ for Windows Version 4.0
 21 <210> SEQ ID NO: 1
 22 <211> LENGTH: 2710
 23 <212> TYPE: DNA
 24 <213> ORGANISM: Trichoderma reesei
 26 <400> SEQUENCE: 1

27	ccacgcgtcc	gagcagtgtc	ctcctcctca	ctgcttcgtc	atgaaggtct	ctcgagtcct	60
28	tgcccttgtc	ctgggggccc	tcctccctgc	ccatgctgcc	ttttcatgga	agaacgtcaa	120
29	gctcggcgcc	ggcggcggtc	tcgtccccgg	catcatcttc	catcccaaga	caaaaggcgt	180
30	agcatatgca	cgaacagata	ttggcgggct	gtaccgcctc	aacgccgacg	actcatggac	240
31	cgccgtcacg	gatgggattg	ctgataatgc	cggctggcac	aactggggca	tcgacgctgt	300
32	tgcgcttgat	cgcgaggacg	atcaaaaggt	gtatgccgca	gtcggcatgt	atacgaacag	360
33	ctgggatccg	agtaatggag	ccatcattcg	ctcgtcagac	cgcggcgcaa	cgtggtcctt	420
34	caccaacttg	cccttcaaag	tcgggggtaa	catgccagga	cgcggagccg	gagagcgtct	480
35	ggctgtcgat	cgggccaaact	ccaacatcat	ctactttggt	gctcgtctag	gaaacggcct	540
36	ctggaagtct	acggacggcg	gcgtgacctt	ttccaaggtc	tcgtcgttca	cggcaactgg	600
37	gacgtacatc	ccagaccgga	gtgattccaa	cggctacaac	agcgacaagc	aaggactcat	660
38	gtgggttacg	ttcgactcaa	ccagcagcac	gaccggggga	gccacgtctc	gtatctttgt	720
39	tggcacggct	gataacatca	ctgcttcagt	ctatgtgagc	acgaatgccg	gctccacgtg	780
40	gagtgtctga	ccggggcagc	cagggaaaata	ctttcctcac	aaggcgaaac	tgacgccagc	840
41	agagaaggcc	ttgtatctga	cctattccga	tggcacaggg	ccgtatgatg	gcacacttgg	900
42	ctcagtgtgg	aggtacgaca	ttgcaggggg	aacttggaaa	gacatcaccc	ctgtctctgg	960
43	atcagatcta	tactttggct	ttggcgccct	tggcctcgat	ttgcaaaagc	caggaaccct	1020
44	tgttgttgct	tctttgaact	cttggtggcc	agatgctcag	ctgtttcggt	cgaccgactc	1080
45	tgggacaaca	tggagcccga	tctgggcgtg	ggcgagctat	ccgactgaga	cctattacta	1140
46	cagcatctca	actcccaaag	caccgtggat	caagaacaac	tttatcgatg	tgacgagcga	1200
47	gtcaccgtcc	gatggtctca	tcaagcgctt	cggctggatg	attgagtctc	tcgagattga	1260
48	cccaaccgac	agcaaccact	ggctctacgg	caccggaatg	acaatctttg	gcggccacga	1320
49	tctcaccaac	tgggacacgc	gccacaatgt	gtcaatccaa	tactggcgag	acggcatcga	1380
50	ggaattctcc	gtccaggacc	tggcctctgc	acccggcgga	agcgagctat	tggccgcagt	1440
51	cggagacgac	aacggcttca	cctttgccag	cagaaacgac	ctcgggacat	cgccgcagac	1500
52	ggtctgggca	acgcccacat	gggccacctc	gacgagcgtc	gactacgccg	ggaactcggt	1560
53	caagagcgtc	gtccgcgctc	gcaacaccgc	cggcacgcaa	caggtggcca	tctcgtccga	1620

RAW SEQUENCE LISTING

DATE: 05/13/2002

PATENT APPLICATION: US/10/026,994

TIME: 09:38:32

Input Set : A:\GC698-SEQLIST.txt

Output Set: N:\CRF3\05102002\J026994.raw

```

54 cggcgggöcgcg acgtggagca tcgactacgc ggccgacacg tccatgaacg gcggcacggg 1680
55 ggcctattcg gccgacggcg acacgatacct ctggtcgacc gcctcgtccg gcgtgcagcg 1740
56 ctöcgagttc cagggcagct ttgcctccgt ctöcagcctg cccgcgggöcg ccgtcatcgc 1800
57 ctöcggacaag aagaccaaca gcgtcttcta cgcöggctcc ggatcgcact ttacgtcag 1860
58 caaggacacc ggcagcagct tcacgcgcgg gcccaagctg ggcagcgcag ggacgatccg 1920
59 ggatatcgct gtcacccöga ccaccgcggg cacgttgat gtctcgaccg acgtcggcat 1980
60 attccgctcc acagactcgg gcacgacctt tggccaagtc tccaccgccc tgaccaacac 2040
61 ctaccagatc gccctgggtg tgggtcagg ctögaactgg aacctgtatg ccttcggcac 2100
62 cggcccgta ggggctcgc tctacgccag tggagacagc ggcgcctcct ggacggacat 2160
63 ccagggtcc cagggttcg gctccatcga cagcaccaag gtcgcgggca gcggcagcac 2220
64 cgcögggcaa gtctacgtg gcaccaacgg ccggggcgtc ttttacgtc agggaaaccgt 2280
65 cggcggcggc acgggcggga cttcctcgtc gaccaagcag agcagcagca gtacctcttc 2340
66 cgcacgctcg agcaccacgc tgaggtcgag cgttgatcc acgaccggg cttcgcagg 2400
67 gacttcgtcg aggaccagct cggcgcggg tccacgggg tcaggggctc ccggtcatta 2460
68 tgctcagtc ggagggattg ggtggacggg gccgacgcag tgtgtggcgc cgtatgtctg 2520
69 ccagaagcag aatgattatt actaccagtg tgtgtgatgc ttgaactgcc aagctcacga 2580
70 ggagagctac ataccöcctag gctcgcagta aagagctcaa gcacccgaag aagcactagt 2640
71 agtagagatc cagtcagata attatccatt tgtttgaatt aaatgatctt ctattgaaaa 2700
72 aaaaaaaaaa 2710

```

74 <210> SEQ ID NO: 2

75 <211> LENGTH: 818

76 <212> TYPE: PRT

77 <213> ORGANISM: Trichoderma reesei

79 <400> SEQUENCE: 2

```

80 Ala Phe Ser Trp Lys Asn Val Lys Leu Gly Gly Gly Gly Gly Phe Val
81 1 5 10 15
82 Pro Gly Ile Ile Phe His Pro Lys Thr Lys Gly Val Ala Tyr Ala Arg
83 20 25 30
84 Thr Asp Ile Gly Gly Leu Tyr Arg Leu Asn Ala Asp Asp Ser Trp Thr
85 35 40 45
86 Ala Val Thr Asp Gly Ile Ala Asp Asn Ala Gly Trp His Asn Trp Gly
87 50 55 60
88 Ile Asp Ala Val Ala Leu Asp Pro Gln Asp Asp Gln Lys Val Tyr Ala
89 65 70 75 80
90 Ala Val Gly Met Tyr Thr Asn Ser Trp Asp Pro Ser Asn Gly Ala Ile
91 85 90 95
92 Ile Arg Ser Ser Asp Arg Gly Ala Thr Trp Ser Phe Thr Asn Leu Pro
93 100 105 110
94 Phe Lys Val Gly Gly Asn Met Pro Gly Arg Gly Ala Gly Glu Arg Leu
95 115 120 125
96 Ala Val Asp Pro Ala Asn Ser Asn Ile Ile Tyr Phe Gly Ala Arg Ser
97 130 135 140
98 Gly Asn Gly Leu Trp Lys Ser Thr Asp Gly Gly Val Thr Phe Ser Lys
99 145 150 155 160
100 Val Ser Ser Phe Thr Ala Thr Gly Thr Tyr Ile Pro Asp Pro Ser Asp
101 165 170 175
102 Ser Asn Gly Tyr Asn Ser Asp Lys Gln Gly Leu Met Trp Val Thr Phe
103 180 185 190
104 Asp Ser Thr Ser Ser Thr Thr Gly Gly Ala Thr Ser Arg Ile Phe Val

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/026,994

DATE: 05/13/2002

TIME: 09:38:32

Input Set : A:\GC698-SEQLIST.txt

Output Set: N:\CRF3\05102002\J026994.raw

```

105          195          200          205
106 Gly Thr Ala Asp Asn Ile Thr Ala Ser Val Tyr Val Ser Thr Asn Ala
107          210          215          220
108 Gly Ser Thr Trp Ser Ala Val Pro Gly Gln Pro Gly Lys Tyr Phe Pro
109 225          230          235          240
110 His Lys Ala Lys Leu Gln Pro Ala Glu Lys Ala Leu Tyr Leu Thr Tyr
111          245          250          255
112 Ser Asp Gly Thr Gly Pro Tyr Asp Gly Thr Leu Gly Ser Val Trp Arg
113          260          265          270
114 Tyr Asp Ile Ala Gly Gly Thr Trp Lys Asp Ile Thr Pro Val Ser Gly
115          275          280          285
116 Ser Asp Leu Tyr Phe Gly Phe Gly Gly Leu Gly Leu Asp Leu Gln Lys
117          290          295          300
118 Pro Gly Thr Leu Val Val Ala Ser Leu Asn Ser Trp Trp Pro Asp Ala
119 305          310          315          320
120 Gln Leu Phe Arg Ser Thr Asp Ser Gly Thr Thr Trp Ser Pro Ile Trp
121          325          330          335
122 Ala Trp Ala Ser Tyr Pro Thr Glu Thr Tyr Tyr Ser Ile Ser Thr
123          340          345          350
124 Pro Lys Ala Pro Trp Ile Lys Asn Phe Ile Asp Val Thr Ser Glu
125          355          360          365
126 Ser Pro Ser Asp Gly Leu Ile Lys Arg Leu Gly Trp Met Ile Glu Ser
127          370          375          380
128 Leu Glu Ile Asp Pro Thr Asp Ser Asn His Trp Leu Tyr Gly Thr Gly
129 385          390          395          400
130 Met Thr Ile Phe Gly Gly His Asp Leu Thr Asn Trp Asp Thr Arg His
131          405          410          415
132 Asn Val Ser Ile Gln Ser Leu Ala Asp Gly Ile Glu Glu Phe Ser Val
133          420          425          430
134 Gln Asp Leu Ala Ser Ala Pro Gly Gly Ser Glu Leu Leu Ala Ala Val
135          435          440          445
136 Gly Asp Asp Asn Gly Phe Thr Phe Ala Ser Arg Asn Asp Leu Gly Thr
137          450          455          460
138 Ser Pro Gln Thr Val Trp Ala Thr Pro Thr Trp Ala Thr Ser Thr Ser
139 465          470          475          480
140 Val Asp Tyr Ala Gly Asn Ser Val Lys Ser Val Val Arg Val Gly Asn
141          485          490          495
142 Thr Ala Gly Thr Gln Val Ala Ile Ser Ser Asp Gly Gly Ala Thr Trp
143          500          505          510
144 Ser Ile Asp Tyr Ala Ala Asp Thr Ser Met Asn Gly Gly Thr Val Ala
145          515          520          525
146 Tyr Ser Ala Asp Gly Asp Thr Ile Leu Trp Ser Thr Ala Ser Ser Gly
147          530          535          540
148 Val Gln Arg Ser Gln Phe Gln Gly Ser Phe Ala Ser Val Ser Ser Leu
149 545          550          555          560
150 Pro Ala Gly Ala Val Ile Ala Ser Asp Lys Lys Thr Asn Ser Val Phe
151          565          570          575
152 Tyr Ala Gly Ser Gly Ser Thr Phe Tyr Val Ser Lys Asp Thr Gly Ser
153          580          585          590

```

RAW SEQUENCE LISTING

DATE: 05/13/2002

PATENT APPLICATION: US/10/026,994

TIME: 09:38:32

Input Set : A:\GC698-SEQLIST.txt

Output Set: N:\CRF3\05102002\J026994.raw

```

154 Ser Phe Thr Arg Gly Pro Lys Leu Gly Ser Ala Gly Thr Ile Arg Asp
155           595           600           605
156 Ile Ala Ala His Pro Thr Thr Ala Gly Thr Leu Tyr Val Ser Thr Asp
157           610           615           620
158 Val Gly Ile Phe Arg Ser Thr Asp Ser Gly Thr Thr Phe Gly Gln Val
159 625           630           635           640
160 Ser Thr Ala Leu Thr Asn Thr Tyr Gln Ile Ala Leu Gly Val Gly Ser
161           645           650           655
162 Gly Ser Asn Trp Asn Leu Tyr Ala Phe Gly Thr Gly Pro Ser Gly Ala
163           660           665           670
164 Arg Leu Tyr Ala Ser Gly Asp Ser Gly Ala Ser Trp Thr Asp Ile Gln
165           675           680           685
166 Gly Ser Gln Gly Phe Gly Ser Ile Asp Ser Thr Lys Val Ala Gly Ser
167           690           695           700
168 Gly Ser Thr Ala Gly Gln Val Tyr Val Gly Thr Asn Gly Arg Gly Val
169 705           710           715           720
170 Phe Tyr Ala Gln Gly Thr Val Gly Gly Gly Thr Gly Gly Thr Ser Ser
171           725           730           735
172 Ser Thr Lys Gln Ser Ser Ser Ser Thr Ser Ser Ala Ser Ser Ser Thr
173           740           745           750
174 Thr Leu Arg Ser Ser Val Val Ser Thr Thr Arg Ala Ser Thr Val Thr
175           755           760           765
176 Ser Ser Arg Thr Ser Ser Ala Ala Gly Pro Thr Gly Ser Gly Val Ala
177           770           775           780
178 Gly His Tyr Ala Gln Cys Gly Gly Ile Gly Trp Thr Gly Pro Thr Gln
179 785           790           795           800
180 Cys Val Ala Pro Tyr Val Cys Gln Lys Gln Asn Asp Tyr Tyr Tyr Gln
181           805           810           815
182 Cys Val

```

185 <210> SEQ ID NO: 3

186 <211> LENGTH: 19

187 <212> TYPE: PRT

188 <213> ORGANISM: Trichoderma reesei

190 <400> SEQUENCE: 3

191 Met Lys Val Ser Arg Val Leu Ala Leu Val Leu Gly Ala Val Ile Pro

192 1 5 10 15

193 Ala His Ala

196 <210> SEQ ID NO: 4

197 <211> LENGTH: 2517

198 <212> TYPE: DNA

199 <213> ORGANISM: Trichoderma reesei

201 <400> SEQUENCE: 4

```

202 atgaaggtct ctgagtcct tgcccttgct ctggggggccg tcatccctgc ccatgctgcc 60
203 ttttcatgga agaacgtcaa gctcggcggc ggcggcggt tcgtccccgg catcatcttc 120
204 catcccaaga caaaggcgt agcatatgca cgaacagata ttggcgggct gtaccgcctc 180
205 aacgcgcgac actcatggac cgccgtcacg gatgggattg ctgataatgc cggctggcac 240
206 aactggggca tcgacgtgt tgcgcttgat ccgcaggacg atcaaaagggt gtatgccgca 300
207 gtcggcatgt atacgaacag ctgggatccg agtaatggag ccatcattcg ctgcgcagac 360
208 cgcggcgcaa cgtggtcctt caccaacttg cccttcaaag tcgggggtaa catgccagga 420

```

RAW SEQUENCE LISTING

DATE: 05/13/2002

PATENT APPLICATION: US/10/026,994

TIME: 09:38:32

Input Set : A:\GC698-SEQLIST.txt

Output Set: N:\CRF3\05102002\J026994.raw

209	cgcgaggaccg	gagagcgtct	ggctgtcgat	ccggccaact	ccaacatcat	ctactttggt	480
210	gctcgctcag	gaaacggcct	ctggaagtct	acggacggcg	gcgtgacctt	ttccaaggtc	540
211	tcgtcgttca	cggcaacttg	gacgtacatc	ccagaccga	gtgattccaa	cggctacaac	600
212	agcgacaagc	aaggactcat	gtgggttacg	ttcgactcaa	ccagcagcac	gaccggggga	660
213	gccacgtctc	gtatctttgt	tggcacggct	gataacatca	ctgcttcagt	ctatgtgagc	720
214	acgaatgccg	gctccacgtg	gagtgtgtga	ccggggcagc	cagggaaata	ctttcctcac	780
215	aaggcgaaac	tgcagccagc	agagaaggcc	ttgtatctga	cctattccga	tggcacaggg	840
216	ccgtatgatg	gcacacttgg	ctcagtgtgg	aggtacgaca	ttgcaggggg	aacttggaaa	900
217	gacatcaccc	ctgtctctgg	atcagatcta	tactttggct	ttggcggcct	tggcctcgat	960
218	ttgcaaaagc	caggaacctt	tgttgttgc	tctttgaact	cttgggtggc	agatgctcag	1020
219	ctgtttcggt	cgaccgactc	tgggacaaca	tggagccga	tctgggcgtg	ggcgagctat	1080
220	ccgactgaga	cctattacta	cagcatctca	actcccaaag	caccgtggat	caagaacaac	1140
221	tttatcgatg	tgcagagcga	gtcaccgctc	gatggtctca	tcaagcgctt	cggctggatg	1200
222	attgagtctc	tcgagattga	cccaaccgac	agcaaccact	ggctctacgg	caccggaatg	1260
223	acaatctttg	gcggccaaga	tctcaccaac	tgggacacgc	gccacaatgt	gtcaatccaa	1320
224	tactggcag	acggcatcga	ggaattctcc	gtccaggacc	tggcctctgc	acccggcgga	1380
225	agcgagctat	tggccgcagt	cggagacgac	aacggcttca	cctttgccag	cagaaacgac	1440
226	ctcgggacat	cgccgcagac	ggtctgggca	acgcccacat	gggccacctc	gacgagcgtc	1500
227	gactacgccg	ggaactcggg	caagagcgtc	gtccgcgtcg	gcaacaccgc	cggcacgcaa	1560
228	caggtggcca	tctcgtccga	cggcggcgcg	acgtggagca	tcgactacgc	ggccgacacg	1620
229	tccatgaacg	gcggcaacgg	ggcctattcg	gccgacggcg	acacgatcct	ctggctgacc	1680
230	gcctcgtccg	gcgtgcagcg	ctcgcagttc	cagggcagct	ttgcctccgt	ctcgagcctg	1740
231	cccgcggggc	ccgtcatcgc	ctcggacaag	aagaccaaca	gcgtcttcta	cgcggctcc	1800
232	ggatcgacct	tttacgtcag	caaggacaac	ggcagcagct	tcacgcgcgg	gccaagctg	1860
233	ggcagcgag	ggacgatccg	ggatatcgct	gtcacccga	ccaccgcggg	cacgttgtat	1920
234	gtctcgaccg	acgtcggcat	attccgctcc	acagactcgg	gcacgacctt	tggccaagtc	1980
235	tccaccgccc	tgaccaaacac	ctaccagatc	gccctgggtg	tgggctcagg	ctcgaactgg	2040
236	aacctgtatg	ccttcggcac	cggcccgtca	ggggctcgcc	tctacgccag	tggagacagc	2100
237	ggcgctcct	ggacggacat	ccagggctcc	cagggcttcg	gctccatcga	cagcaccaag	2160
238	gtcgccggca	gcggcagcac	cgcggggcaa	gtctacgtgg	gcaccaacgg	ccggggcgctc	2220
239	ttttacgctc	agggaaaccgt	cggcgggcgg	acggggcgga	cttcctcgtc	gaccaagcag	2280
240	agcagcagca	gtacctcttc	cgcagctcgc	agcaccacgc	tgaggtcgag	cgttgtatcc	2340
241	acgacccggg	cttcgacggg	gaattcgctg	aggaccagct	cggccgcggg	tcccacgggg	2400
242	tcaggggtcg	ccggtcatta	tgtcagtgcc	ggagggattg	ggtggacggg	gccgacgcag	2460
243	tgtgtggcgc	cgtatgtctg	ccagaagcag	aatgattatt	actaccagtg	tgtgtga	2517

VERIFICATION SUMMARY

DATE: 05/13/2002

PATENT APPLICATION: US/10/026,994

TIME: 09:38:33

Input Set : A:\GC698-SEQLIST.txt

Output Set: N:\CRF3\05102002\J026994.raw

L:15 M:271 C: Current Filing Date differs, Replaced Current Filing Date